

## RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.

Application Serial Number: 10765, 120 A  
Source: FWO  
Date Processed by STIC: 11-3-04

# ***ENTERED***

**CRF Errors Edited by the STIC Systems  
Branch**

Serial Number: 10765, 120A

CRF Edit Date: 11/3/04  
Edited by: KL

\_\_\_ Realigned nucleic acid/amino acid numbers/text in cases where the sequence text "wrapped" to the next line

\_\_\_ Corrected the SEQ ID NO. Sequence numbers edited were:

\_\_\_\_\_

\_\_\_ Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited:

\_\_\_\_\_

\_\_\_ Deleted: \_\_\_ invalid beginning/end-of-file text ; \_\_\_ page numbers

\_\_\_ Inserted mandatory headings/numeric identifiers, specifically:

\_\_\_\_\_

\_\_\_ Moved responses to same line as heading/numeric identifier, specifically:

\_\_\_\_\_

✓ Other:

For Seq ID # 35, inserted indentation,  
Also corrected Seq ID # 38, 42107  
to 38.



IFWO

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/765,120A

DATE: 11/03/2004

TIME: 09:41:46

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw

```

1 <110> APPLICANT: Benner, Steven Albert
3 <120> TITLE OF INVENTION: Evolution-Based Functional Genomics
5 <130> FILE REFERENCE: file reference 10-765120
7 <140> CURRENT APPLICATION NUMBER: 10/765,120A
8 <141> CURRENT FILING DATE: 2004-01-28
10 <160> NUMBER OF SEQ ID NOS: 38
12 <170> SOFTWARE: MacIntosh OS 10.3 Microsoft Word v. 2003
14 <210> SEQ ID NO: 1
15 <211> LENGTH: 486
16 <212> TYPE: PRT
17 <213> ORGANISM: Tilapia nilotica
19 <400> SEQUENCE: 1
20 Met Val Leu Glu Met Leu Asn Pro Met His Tyr Lys Val Thr Ser
21                               5                10                15
22 Met Val Ser Glu Val Val Pro Phe Ala Ser Ile Ala Val Leu Leu
23                               20                25                30
24 Leu Thr Gly Phe Leu Leu Leu Val Trp Asn Tyr Lys Asn Thr Ser
25                               35                40                45
26 Ser Ile Pro Gly Pro Gly Tyr Phe Leu Gly Ile Gly Pro Leu Ile
27                               50                55                60
28 Ser Tyr Leu Arg Phe Leu Trp Met Gly Ile Gly Ser Ala Cys Asn
29                               65                70                75
30 Tyr Tyr Asn Lys Thr Tyr Gly Glu Phe Ile Arg Val Trp Ile Gly
31                               80                85                90
32 Gly Glu Glu Thr Leu Ile Ile Ser Lys Ser Ser Ser Val Phe His
33                               95               100               105
34 Val Met Lys His Ser His Tyr Thr Ser Arg Phe Gly Ser Lys Pro
35                               110              115              120
36 Gly Leu Gln Phe Ile Gly Met His Glu Lys Gly Ile Ile Phe Asn
37                               125              130              135
38 Asn Asn Pro Val Leu Trp Lys Ala Val Arg Thr Tyr Phe Met Lys
39                               140              145              150
40 Ala Leu Ser Gly Pro Gly Leu Val Arg Met Val Thr Val Cys Ala
41                               155              160              165
42 Asp Ser Ile Thr Lys His Leu Asp Lys Leu Glu Glu Val Arg Asn
43                               170              175              180
44 Asp Leu Gly Tyr Val Asp Val Leu Thr Leu Met Arg Arg Ile Met
45                               185              190              195
46 Leu Asp Thr Ser Asn Asn Leu Phe Leu Gly Ile Pro Leu Asp Glu
47                               200              205              210
48 Lys Ala Ile Val Cys Lys Ile Gln Gly Tyr Phe Asp Ala Trp Gln
49                               215              220              225
50 Ala Leu Leu Leu Lys Pro Asp Ile Phe Phe Lys Ile Pro Trp Leu

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/765,120A

DATE: 11/03/2004

TIME: 09:41:46

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw

```

51          230          235          240
52 Tyr Arg Lys Tyr Glu Lys Ser Val Lys Asp Leu Lys Glu Asp Met
53          245          250          255
54 Glu Ile Leu Ile Glu Lys Lys Arg Arg Arg Ile Phe Thr Ala Glu
55          260          265          270
56 Lys Leu Glu Asp Cys Met Asp Phe Ala Thr Glu Leu Ile Leu Ala
57          275          280          285
58 Glu Lys Arg Gly Glu Leu Thr Lys Glu Asn Val Asn Gln Cys Ile
59          290          295          300
60 Leu Glu Met Leu Ile Ala Ala Pro Asp Thr Met Ser Val Thr Val
61          305          310          315
62 Phe Phe Met Leu Phe Leu Ile Ala Lys His Pro Gln Val Glu Glu
63          320          325          330
64 Glu Leu Met Lys Glu Ile Gln Thr Val Val Gly Glu Arg Asp Ile
65          335          340          345
66 Arg Asn Asp Asp Met Gln Lys Leu Glu Val Val Glu Asn Phe Ile
67          350          355          360
68 Tyr Glu Ser Met Arg Tyr Gln Pro Val Val Asp Leu Val Met Arg
69          365          370          375
70 Lys Ala Leu Glu Asp Asp Val Ile Asp Gly Tyr Pro Val Lys Lys
71          380          385          390
72 Gly Thr Asn Ile Ile Leu Asn Ile Gly Arg Met His Arg Leu Glu
73          395          400          405
74 Phe Phe Pro Lys Pro Asn Glu Phe Thr Leu Glu Asn Phe Ala Lys
75          410          415          420
76 Asn Val Pro Tyr Arg Tyr Phe Gln Pro Phe Gly Phe Gly Pro Arg
77          425          430          435
78 Ala Cys Ala Gly Lys Tyr Ile Ala Met Val Met Met Lys Val Thr
79          440          445          450
80 Leu Val Ile Leu Leu Arg Arg Phe Gln Val Gln Thr Pro Gln Asp
81          455          460          465
82 Arg Cys Val Glu Lys Met Gln Lys Lys Asn Asp Leu Ser Leu His
83          470          475          480
84 Pro Asp Glu Thr Ser Gly
85          485
87 <210> SEQ ID NO: 2
88 <211> LENGTH: 486
89 <212> TYPE: PRT
90 <213> ORGANISM: Oryzias latipes
92 <400> SEQUENCE: 2
93 Met Phe Leu Glu Met Leu Asn Pro Met Gln Tyr Asn Val Thr Ile
94          5          10          15
95 Met Val Pro Glu Thr Val Thr Val Ser Ala Met Pro Leu Leu Leu
96          20          25          30
97 Ile Met Gly Leu Leu Leu Leu Ile Trp Asn Cys Glu Ser Ser Ser
98          35          40          45
99 Ser Ile Pro Gly Pro Gly Tyr Cys Leu Gly Ile Gly Pro Leu Ile
100          50          55          60
101 Ser His Gly Arg Phe Leu Trp Met Gly Ile Gly Ser Ala Cys Asn

```

## RAW SEQUENCE LISTING

DATE: 11/03/2004

PATENT APPLICATION: US/10/765,120A

TIME: 09:41:46

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw

102		65		70		75
103	Tyr Tyr Asn Lys Met Tyr Gly Glu Phe Met Arg Val Trp Ile Ser					
104		80		85		90
105	Gly Glu Glu Thr Leu Ile Ile Ser Lys Ser Ser Ser Met Phe His					
106		95		100		105
107	Val Met Lys His Ser His Tyr Ile Ser Arg Phe Gly Ser Lys Arg					
108		110		115		120
109	Gly Leu Gln Cys Ile Gly Met His Glu Asn Gly Ile Ile Phe Asn					
110		125		130		135
111	Asn Asn Pro Ser Leu Trp Arg Thr Ile Arg Pro Phe Phe Met Lys					
112		140		145		150
113	Ala Leu Thr Gly Pro Gly Leu Val Arg Met Val Glu Val Cys Val					
114		155		160		165
115	Glu Ser Ile Lys Gln His Leu Asp Arg Leu Gly Glu Val Thr Asp					
116		170		175		180
117	Thr Ser Gly Tyr Val Asp Val Leu Thr Leu Met Arg His Ile Met					
118		185		190		195
119	Leu Asp Thr Ser Asn Met Leu Phe Leu Gly Ile Pro Leu Asp Glu					
120		200		205		210
121	Ser Ala Ile Val Lys Lys Ile Gln Gly Tyr Phe Asn Ala Trp Gln					
122		215		220		225
123	Ala Leu Leu Ile Lys Pro Asn Ile Phe Phe Lys Ile Ser Trp Leu					
124		230		235		240
125	Tyr Arg Lys Tyr Glu Arg Ser Val Lys Asp Leu Lys Asp Glu Ile					
126		245		250		255
127	Ala Val Leu Val Glu Lys Lys Arg His Lys Val Ser Thr Ala Glu					
128		260		265		270
129	Lys Leu Glu Asp Cys Met Asp Phe Ala Thr Asp Leu Ile Phe Ala					
130		275		280		285
131	Glu Arg Arg Gly Asp Leu Thr Lys Glu Asn Val Asn Gln Cys Ile					
132		290		295		300
133	Leu Glu Met Leu Ile Ala Ala Pro Asp Thr Met Ser Val Thr Leu					
134		305		310		315
135	Tyr Phe Met Leu Leu Leu Val Ala Glu Tyr Pro Glu Val Glu Ala					
136		320		325		330
137	Ala Ile Leu Lys Glu Ile His Thr Val Val Gly Asp Arg Asp Ile					
138		335		340		345
139	Lys Ile Glu Asp Ile Gln Asn Leu Lys Val Val Glu Asn Phe Ile					
140		350		355		360
141	Asn Glu Ser Met Arg Tyr Gln Pro Val Val Asp Leu Val Met Arg					
142		365		370		375
143	Arg Ala Leu Glu Asp Asp Val Ile Asp Gly Tyr Pro Val Lys Lys					
144		380		385		390
145	Gly Thr Asn Ile Ile Leu Asn Ile Gly Arg Met His Arg Leu Glu					
146		395		400		405
147	Tyr Phe Pro Lys Pro Asn Glu Phe Thr Leu Glu Asn Phe Glu Lys					
148		410		415		420
149	Asn Val Pro Tyr Arg Tyr Phe Gln Pro Phe Gly Phe Gly Pro Arg					
150		425		430		435

## RAW SEQUENCE LISTING

DATE: 11/03/2004

PATENT APPLICATION: US/10/765,120A

TIME: 09:41:46

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw

```

151 Gly Cys Ala Gly Lys Tyr Ile Ala Met Val Met Met Lys Val Val
152                               440                      445          450
153 Leu Val Thr Leu Leu Arg Arg Phe Gln Val Lys Thr Leu Gln Lys
154                               455                      460          465
155 Arg Cys Ile Glu Asn Ile Pro Lys Lys Asn Asp Leu Ser Leu His
156                               470                      475          480
157 Pro Asn Glu Asp Arg His
158                               485
160 <210> SEQ ID NO: 3
161 <211> LENGTH: 486
162 <212> TYPE: PRT
163 <213> ORGANISM: Danio rerio
165 <400> SEQUENCE: 3
166 Met Ile Leu Glu Met Leu Asn Pro Met His Tyr Asn Leu Thr Ser
167                               5                      10              15
168 Met Val Pro Glu Val Met Pro Val Ala Thr Leu Pro Ile Leu Leu
169                               20                      25              30
170 Leu Thr Gly Phe Leu Phe Phe Val Trp Asn His Glu Glu Thr Ser
171                               35                      40              45
172 Ser Ile Pro Gly Pro Gly Tyr Cys Met Gly Ile Gly Pro Leu Ile
173                               50                      55              60
174 Ser His Leu Arg Phe Leu Trp Met Gly Leu Gly Ser Ala Cys Asn
175                               65                      70              75
176 Tyr Tyr Asn Lys Met Tyr Gly Glu Phe Val Arg Val Trp Ile Ser
177                               80                      85              90
178 Gly Glu Glu Thr Leu Val Ile Ser Lys Ser Ser Ser Thr Phe His
179                               95                      100             105
180 Ile Met Lys His Asp His Tyr Ser Ser Arg Phe Gly Ser Thr Phe
181                               110                     115             120
182 Gly Leu Gln Tyr Met Gly Met His Glu Asn Gly Val Ile Phe Asn
183                               125                     130             135
184 Asn Asn Pro Ala Val Trp Lys Ala Leu Arg Pro Phe Phe Val Lys
185                               140                     145             150
186 Ala Leu Ser Gly Pro Ser Leu Ala Arg Met Val Thr Val Cys Val
187                               155                     160             165
188 Glu Ser Val Asn Asn His Leu Asp Arg Leu Asp Glu Val Thr Asn
189                               170                     175             180
190 Ala Leu Gly His Val Asn Val Leu Thr Leu Met Arg Arg Thr Met
191                               185                     190             195
192 Leu Asp Ala Ser Asn Thr Leu Phe Leu Arg Ile Pro Leu Asp Glu
193                               200                     205             210
194 Lys Asn Ile Val Leu Lys Ile Gln Gly Tyr Phe Asp Ala Trp Gln
195                               215                     220             225
196 Ala Leu Leu Ile Lys Pro Asn Ile Phe Phe Lys Ile Ser Trp Leu
197                               230                     235             240
198 Ser Arg Lys His Gln Lys Ser Ile Lys Glu Leu Arg Asp Ala Val
199                               245                     250             255
200 Gly Ile Leu Ala Glu Glu Lys Arg His Arg Ile Phe Thr Ala Glu
201                               260                     265             270

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/765,120A

DATE: 11/03/2004

TIME: 09:41:46

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw

```

202 Lys Leu Glu Asp His Val Asp Phe Ala Thr Asp Leu Ile Leu Ala
203          275          280          285
204 Glu Lys Arg Gly Glu Leu Thr Lys Glu Asn Val Asn Gln Cys Ile
205          290          295          300
206 Leu Glu Met Met Ile Ala Ala Pro Asp Thr Leu Ser Val Thr Val
207          305          310          315
208 Phe Phe Met Leu Cys Leu Ile Ala Gln His Pro Lys Val Glu Glu
209          320          325          330
210 Ala Leu Met Lys Glu Ile Gln Thr Val Leu Gly Glu Arg Asp Leu
211          335          340          345
212 Lys Asn Asp Asp Met Gln Lys Leu Lys Val Met Glu Asn Phe Ile
213          350          355          360
214 Asn Glu Ser Met Arg Tyr Gln Pro Val Val Asp Ile Val Met Arg
215          365          370          375
216 Lys Ala Leu Glu Asp Asp Val Ile Asp Gly Tyr Pro Val Lys Lys
217          380          385          390
218 Gly Thr Asn Ile Ile Leu Asn Ile Gly Arg Met His Lys Leu Glu
219          395          400          405
220 Phe Phe Pro Lys Pro Asn Glu Phe Thr Leu Glu Asn Phe Glu Lys
221          410          415          420
222 Asn Val Pro Tyr Arg Tyr Phe Gln Pro Phe Gly Phe Gly Pro Arg
223          425          430          435
224 Ser Cys Ala Gly Lys Phe Ile Ala Met Val Met Met Lys Val Met
225          440          445          450
226 Leu Val Ser Leu Leu Arg Arg Phe His Val Lys Thr Leu Gln Gly
227          455          460          465
228 Asn Cys Leu Glu Asn Met Gln Lys Thr Asn Asp Leu Ala Leu His
229          470          475          480
230 Pro Asp Glu Ser Arg Ser
231          485
233 <210> SEQ ID NO: 4
234 <211> LENGTH: 487
235 <212> TYPE: PRT
236 <213> ORGANISM: Carassius auratus
238 <400> SEQUENCE: 4
239 Val Leu Glu Leu Leu Met Gln Gly Ala His Asn Ser Ser Tyr Gly
240          5          10          15
241 Ala Gln Asp Asn Val Cys Gly Ala Met Ala Thr Leu Leu Leu Leu
242          20          25          30
243 Leu Leu Cys Leu Leu Leu Ala Ile Arg His His Trp Thr Glu Lys
244          35          40          45
245 Asp His Val Pro Gly Pro Cys Phe Leu Leu Gly Leu Gly Pro Leu
246          50          55          60
247 Leu Ser Tyr Cys Arg Leu Ile Trp Ser Gly Ile Gly Thr Ala Ser
248          65          70          75
249 Asn Tyr Tyr Asn Ser Lys Tyr Gly Asp Ile Val Arg Val Trp Ile
250          80          85          90
251 Asn Gly Glu Glu Thr Leu Ile Leu Ser Arg Ser Ser Ala Val Tyr
252          95          100          105

```

VERIFICATION SUMMARY

DATE: 11/03/2004

PATENT APPLICATION: US/10/765,120A

TIME: 09:41:47

Input Set : A:\pto.kd.txt

Output Set: N:\CRF4\11032004\J765120A.raw





IFWO

## RAW SEQUENCE LISTING

DATE: 10/27/2004

PATENT APPLICATION: US/10/765,120A

TIME: 10:19:47

Input Set : A:\pto.lm.txt

Output Set: N:\CRF4\10272004\J765120A.raw

1 <110> APPLICANT: Benner, Steven Albert  
 3 <120> TITLE OF INVENTION: Evolution-Based Functional Genomics  
 5 <130> FILE REFERENCE: file reference 10-765120  
 7 <140> CURRENT APPLICATION NUMBER: 10/765,120A  
 8 <141> CURRENT FILING DATE: 2004-01-28  
 10 <160> NUMBER OF SEQ ID NOS: 38  
 12 <170> SOFTWARE: MacIntosh OS 10.3 Microsoft Word v. 2003

## ERRORED SEQUENCES

1488 &lt;210&gt; SEQ ID NO: 35

1489 &lt;211&gt; LENGTH: 84

1490 &lt;212&gt; TYPE: DNA

1491 &lt;213&gt; ORGANISM: Sus scrofa

1493 &lt;400&gt; SEQUENCE: 35

W--&gt; 1494 caatcattac acgtgccgat ttggcagcaa acttgggttg gaatgcattg gcatgcatga 60 aaaagggcatc

E--&gt; 1495 atgtttaaca ataa 84

E--&gt; 1515 &lt;210&gt; SEQ ID NO: 37-38

1516 &lt;211&gt; LENGTH: 84

1517 &lt;212&gt; TYPE: DNA

1518 &lt;213&gt; ORGANISM: White lipped peccary

E--&gt; 1520 &lt;400&gt; SEQUENCE: 38

1521 cagtcactac acatcccgat tcggcagcaa acctgggttg cagttcattg gaatgcatga 60

1522 gaaagggcatc atatttaaca acaa 84

Does Not Comply  
 Corrected Diskette Needed

(ps.1)

← moved to  
 next  
 line.

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/765,120A

DATE: 10/27/2004  
TIME: 10:19:48

Input Set : A:\pto.lm.txt  
Output Set: N:\CRF4\10272004\J765120A.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:35; Line(s) 1494

## VERIFICATION SUMMARY

DATE: 10/27/2004

PATENT APPLICATION: US/10/765,120A

TIME: 10:19:48

Input Set : A:\pto.lm.txt

Output Set: N:\CRF4\10272004\J765120A.raw

L:1494 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:8

L:1495 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:35 ✓

L:1495 M:254 E: No. of Bases conflict, LENGTH:Input:84 Counted:14 SEQ:35 ✓

L:1495 M:252 E: No. of Seq. differs, &lt;211&gt; LENGTH:Input:84 Found:14 SEQ:35 ✓

L:1515 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQ ID NO:37 ✓

L:1520 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:37 differs:38 ✓